

Tecnologías Digitales

variación y cambio en el español de América

I Congreso de la Delegación Argentina de la Asociación de Lingüística y Filología de América Latina (ALFAL) y V Jornadas Internacionales de Filología Hispánica

ORGANIZAÇÃO DE UM BANCO DE DADOS A PARTIR DE JORNAIS DO TRIÂNGULO MINEIRO: CONTRIBUIÇÕES PARA O RESGATE DA SÓCIO-HISTÓRIA DO PORTUGUÊS BRASILEIRO

Juliana Bertucci Barbosa (UFTM, Uberaba / GEVAR / CNPq)
juliana@ielachs.uftm.edu.br

Talita de Cássia Marine (UFU / GEVAR)
talitamarine@ileel.ufu.br

Área temática: *Sociolinguística*

Resumo

O uso de *corpus* favorece as pesquisas linguísticas, sobretudo as de variação e mudança linguística (Weinreich, Labov, Herzog, 2006; Labov, 1994, 2001, 2008), na medida em que proporciona a realização de descrições linguísticas de base empírica, fundamentadas em contextos de usos reais da língua. A fim de recuperar aspectos de sincronias passadas e atuais do Português Mineiro da região do Triângulo (Minas Gerais/Brasil), entendemos como uma importante fonte para pesquisas de cunho variacionista, o jornal, dada a diversidade de gêneros textuais que abarca, bem como ao fato de se mostrar como um veículo de expressão de costumes, cultura, comportamentos, valores, enfim, da história de uma dada sociedade, num dado tempo. Com base nessas observações, apresentaremos os principais aspectos metodológicos que estamos aplicando para a organização de um banco de dados composto por jornais de Uberaba - *Lavoura e Comércio*, *Correio Católico* e *A Gazeta* – e Uberlândia - *Correio de Uberlândia*. Cabe destacar que a digitalização do *Lavoura e Comércio*, por exemplo, fundado em 1889, já foi iniciada e, parte desse jornal já está sendo utilizada como *corpus* para pesquisas linguísticas, apresentando muitas peculiaridades da língua escrita da época. Já o jornal *Correio de Uberlândia*, fundado em 1938 e publicado sem interrupções até hoje, possui todos os exemplares digitalizados e disponíveis no Arquivo Público da cidade de Uberlândia. Entretanto, para facilitar o processo de pesquisas linguísticas com tais exemplares, todo o material será digitado em formato de arquivo “Word”. Posteriormente, pretende-se refinar a organização deste material, dividindo-o por seções e/ou cadernos, bem como por gêneros textuais. Acreditamos que com a montagem desse *corpus*, possibilitaremos a realização de pesquisas que contribuam para o levantamento de características do Português Mineiro do Triângulo e, conseqüentemente, para a construção da identidade do Português Brasileiro. (Apoio CNPq / Processo 476810/2010-8)

Palavras-chave: Gêneros textuais - Variação e mudança linguística - Banco de dados

Identities dinámicas: variación y cambio en el español de América

I Congreso de la Delegación Argentina de la Asociación de Lingüística y Filología de América Latina (ALFAL) y V Jornadas Internacionales de Filología Hispánica

Introdução

“Todo lingüista reconhece que a língua é um fato social, mas nem todos dão a mesma ênfase a esse fato” (Labov 2008: 302)

O uso de *corpus* favorece as pesquisas linguísticas, sobretudo, as de variação e mudança linguística, na medida em que proporciona a realização de descrições linguísticas de base empírica, fundamentadas em contextos de usos reais da língua. A fim de recuperar aspectos de sincronias passadas e atuais do Português Mineiro da região do Triângulo (Minas Gerais, Brasil), entendemos como uma importante fonte para pesquisas de cunho variacionista, o jornal.

Com base nessas observações, apresentaremos, neste artigo, o projeto desenvolvido no âmbito do GEVAR (Grupo de Pesquisas Variacionistas): a constituição de um banco de dados composto por jornais de Uberaba e Uberlândia. Acreditamos que com a montagem desse *corpus*, possibilitaremos a realização de pesquisas que contribuam para o levantamento de características do Português Mineiro do Triângulo e, conseqüentemente, para a construção da identidade do Português Brasileiro.

Pesquisa linguística pautada em *corpus*

Estudos embasados pelos pressupostos teórico-metodológicos do modelo da Variação e Mudança Linguísticas (Weinreich, Labov, Herzog 2006; Labov 1994, 2001, 2008), concebem a língua como uma realidade dinâmica e multifacetada, em constante “fazimento”, tal como propunha Coseriu (1979), e que, portanto, é entendida como um conjunto de variedades resultantes

[...] das peculiaridades das experiências históricas e socioculturais do grupo que a usa: como ele se constitui, como é sua posição na estrutura socioeconômica, como ele se organiza socialmente, quais seus valores e visão de mundo, quais suas possibilidades de acesso à escola, aos meios de informação, e assim por diante (Faraco 2005: 32).

Por isso, a heterogeneidade é a preocupação central da Sociolinguística, visto que essa teoria tem como objeto de estudo a diversidade linguística e o uso da língua no seu contexto social, procurando o entendimento dos mecanismos de variação e

Identities dinámicas: variación y cambio en el español de América

I Congreso de la Delegación Argentina de la Asociación de Lingüística y Filología de América Latina (ALFAL) y V Jornadas Internacionales de Filología Hispánica

mudança, do ponto de vista intra- e extralinguístico. Logo, “a língua não pode mais ser definida como um espaço lógico e abstrato de possibilidades, pré-existente à ação, e que a ação apenas atualizaria” (Mondada 1998: 132), afinal, como destaca Marcuschi (2007: 96), “a língua não pré-existe; ela se dá emergencialmente nas situações concretas de uso”.

Diante disso, embora as culturas que operam com a escrita desenvolvam um padrão de língua codificado em gramáticas, cultivado por grande parcela dos letrados e ensinado pelas escolas, adquirindo, assim, um estatuto de estabilidade e permanência maior do que as outras variedades da língua, isso tudo se constitui apenas como um “refreador temporário de mudanças” (Cf. Faraco 2005: 15), mas não como um inibidor, uma vez que a variação e a mudança são inerentes à língua. Entretanto, essa mutação não lhes tira seu caráter sistêmico, ou seja, “as línguas mudam, mas continuam organizadas e oferecendo a seus falantes os recursos necessários para a circulação dos significados” (Faraco 2005:14). Justamente por isso, os falantes, em geral, não têm muita consciência de que a língua está mudando, uma vez que as mudanças ocorrem de maneira bastante lenta, atingindo sempre partes e não o todo da língua (Cf. Faraco 2005).

Segundo a perspectiva variacionista dos fenômenos da linguagem, os processos de variação se iniciam nas interações orais, cotidianas e coloquiais dos falantes de uma mesma comunidade, desenvolvendo-se a partir da correlação entre as formas alternantes e fatores tanto de natureza linguística, como extralinguística. Assim, a mudança nasce *na* e *da* variação em situações de fala menos formais, uma vez que nessas situações o falante tende a se ver livre de “pressões normativas”, refletindo sua linguagem usual – o vernáculo, tal como definido por Labov (2008).

Diante disso, embora estudos de fenômenos de variação e mudança linguísticas, em uma perspectiva de longo prazo, também possam utilizar como *corpus* textos de língua falada como um ponto de partida, vão naturalmente exigir a ampliação do estudo para momentos anteriores na história da língua, período este que não é possível recorrer à análise da língua falada. Fica-se, assim, destinado para os estudos diacrônicos, o estudo dos fenômenos da variação e mudança em textos escritos.

Identities dinámicas: variación y cambio en el español de América

I Congreso de la Delegación Argentina de la Asociación de Lingüística y Filología de América Latina (ALFAL) y V Jornadas Internacionales de Filología Hispánica

Seguindo parcialmente essa tradição, acreditamos que o uso de *corpus* favorece a realização de descrições linguísticas de base empírica e permite a reflexão sobre questões teóricas fundamentadas em usos reais da língua (Cf. Barbosa e Marine (2007). Entretanto, a essa noção, acrescentamos a de que tanto em pesquisas sincrônicas como em diacrônicas, o relevante não é, necessariamente, a modalidade (oral ou escrita) escolhida, mas, sim, o tipo e o gênero textual que se utiliza para uma dada observação/descrição linguística. Nesse sentido, além de recursos tecnológicos para construirmos e/ou utilizarmos um dado *corpus*, devemos também conhecer o tipo e o gênero textual com que estamos trabalhando, já que cada um deles apresenta peculiaridades que implicam não apenas em aspectos estruturais caracterizadores, mas também em tipos de tema mais recorrentes, grau de formalidade – fator este, cuja observação tem se mostrado fundamental aos estudos variacionistas -, intenção, entre outros. E, o conhecimento dessas características conduz o pesquisador a um caminho mais adequado no tratamento do texto utilizado como *corpus*, seja este oral ou escrito, o que, por sua vez, possibilitará uma análise mais pertinente do objeto de estudo do linguista.

Assim, de acordo com Marine (2009), aplicar o modelo teórico-metodológico laboviano em textos escritos deve ser incentivado, pois se fenômenos linguísticos em variação puderem ser observados na escrita, isso se constituirá como forte indício de um estágio avançado de variação, quiçá de mudança. Afinal, como ressalta Faraco (2005: 26), as mudanças linguísticas

[...] costumam se desencadear na fala informal de grupos socioeconômicos intermediários; avançam pela fala informal de grupos mais altos na estrutura socioeconômica; chegam a situações formais de fala e só então começam a ocorrer na escrita (Faraco 2005: 26).

Todavia, um dos obstáculos para quem realiza pesquisas sobre épocas passadas da língua é o da formação de um *corpus* que contemple uma documentação representativa da época desejada (Labov 1972). Embora o pesquisador procure, com rigor filológico, linguístico e histórico escolher textos que sejam testemunhos fiéis de uma época, muitas vezes isso não é possível.

Identities dynamics: variation and change in the Spanish of América

I Congreso de la Delegación Argentina de la Asociación de Lingüística y Filología de América Latina (ALFAL) y V Jornadas Internacionales de Filología Hispánica

Pensando em tais questões, na tentativa de recuperar aspectos de sincronias passadas do Português Mineiro de Uberaba (MG) e Uberlândia (MG), entendemos como uma importante fonte para pesquisas de cunho variacionista, a chamada “língua/linguagem da imprensa”. Nesse sentido, a montagem de um *corpus* a partir de textos jornalísticos, mostra-se bastante interessante, uma vez que permite, por exemplo, a realização de estudos acerca de fenômenos em variação linguística, precursores, como se sabe, de reanálises gramaticais.

Além disso, os jornais também são um excelente material para a recuperação de aspectos sociais, históricos, econômicos, políticos, científicos, psicológicos, culturais - entre tantos outros - de uma época. Com base nessas observações, pretendemos apresentar a organização de um banco de dados composto dos seguintes jornais de Uberaba: “Lavoura e Comércio”, “Correio Católico”, “A Gazeta” e “Jornal de Uberaba”, bem como o “Correio de Uberlândia”, jornal uberlandense fundado no final da década de 1930 e que circula até hoje, de maneira ininterrupta. Cabe ressaltar que, com a montagem desse *corpus*, buscamos possibilitar a realização de pesquisas que contribuam para o levantamento de características do Português Mineiro de Uberaba, Uberlândia e região e, conseqüentemente, para a construção da história do Português Brasileiro.

Resgate de documentos antigos, montagem de corpora e memória de um povo

Partindo das concepções descritas acima, focalizamos neste artigo, como já mencionado, discussões que envolvem a montagem de banco de dados composto pelo suporte “jornal”. Entretanto, antes de apresentarmos a metodologia utilizada para montagem desse *corpus*, é relevante destacar que sempre que nos referirmos à “memória”, estamos focando – no contexto histórico atual, em que estão presentes os avanços tecnológicos – o registro de um conjunto de informações, quais sejam: os documentos e representações que podem ser consultados, servindo de memória social ou memória de longo prazo. Assim, quando estamos montando um *corpus* para pesquisas linguísticas de sincronias passadas, também estamos colaborando para a

Identities dynamics: variation and change in the Spanish of América

I Congreso de la Delegación Argentina de la Asociación de Lingüística y Filología de América Latina (ALFAL) y V Jornadas Internacionales de Filología Hispánica

preservação do patrimônio cultural brasileiro, isto é, para a preservação da memória cultural de um povo.

Cabe lembrar que na a Legislação Ordinária do Tombamento (lei 25, de 30.11.37), no capítulo 1, artigo 1, encontramos para “Patrimônio Histórico e Artístico Nacional”, a seguinte definição: “conjunto de bens móveis e imóveis existentes no país e cuja conservação seja do interesse público, quer por sua vinculação a fatos memoráveis da história do Brasil, quer por seu excepcional valor arqueológico ou etnológico, bibliográfico ou artístico”. Acrescentando a essa definição, temos ainda a nossa Constituição de 1988 que estabelece, no artigo 216, que:

constituem patrimônio cultural brasileiro os bens de natureza material e imaterial, tomados individualmente ou em conjunto, portadores de referência à identidade, à ação, à memória dos diferentes grupos formadores da sociedade brasileira, nos quais se incluem: I - as formas de expressão; II - os modos de criar, fazer e viver; [...].

Com base nessas definições de patrimônio cultural e, principalmente, da ideia de preservar “formas de expressão”, mencionada na Constituição de 1988, acreditamos que o jornal constitui um rico patrimônio cultural para a humanidade e, portanto, deve ser conservado.

Além disso, no suporte “jornal” circulam diferentes gêneros textuais, o que nos permite trabalhar com a noção de estilo/grau de formalidade e realizar estudos de fenômenos em variação linguística, precursores de reanálises gramaticais.

A montagem do banco de dados composto por jornais de Uberaba e Uberlândia: final do século XIX e começo do século XX

Inicialmente, nosso trabalho busca a digitalização de jornais do final do século XIX e da primeira metade do XX, publicados na cidade de Uberaba e Uberlândia. É importante destacar que parte do material encontra-se no Arquivo Público dessas cidades, apenas organizado em pastas. Muitos deles encontram-se em fase avançada de deterioração e não podem ser consultados pela comunidade. Outros, pertencem a colecionadores e, se não forem digitalizados (com as devidas autorizações), continuarão restritos aos olhares de um único indivíduo. Sendo assim, por meio da

Identities dinámicas: variación y cambio en el español de América

I Congreso de la Delegación Argentina de la Asociación de Lingüística y Filología de América Latina (ALFAL) y V Jornadas Internacionales de Filología Hispánica

digitalização desses jornais, buscamos organizar um arquivo digital que possa preservar importantes informações que com o decorrer do tempo podem se perder para sempre.

Assim, partindo das discussões já expostas neste artigo e visando colaborar com o resgate da Língua, cultura, história e sociedade de Uberaba e Uberlândia, em 2011 iniciamos a montagem do arquivo digitalizado, de acordo com as seguintes etapas de execução do trabalho:

- a) **pesquisa de material e revisão bibliográfica:** foram realizadas consultas a bibliotecas – da Universidade do Triângulo Mineiro (UFTM), da Universidade Federal de Uberlândia (UFU) e também nas Bibliotecas Municipais - , ao Arquivo Público de Uberaba e Uberlândia, além de consultas a alguns moradores que possuem coleções particulares (exemplares de jornais antigos em suas casas), visitas a sebos e consultas a sites na Internet;
- b) **seleção de jornais:** escolhemos o material a ser digitalizado, priorizando os mais antigos, deteriorados devido ao “passar do tempo”, focalizando, especialmente, em dois jornais, dada a sua circulação ininterrupta: “Lavoura e Comercio”, de Uberaba, e “Correio de Uberlândia”, de Uberlândia;
- c) **processo de digitalização:** cada página do jornal é transformada em imagem (formato JPEG) por meio de câmeras fotográficas profissionais, financiadas pelo projeto, e/ou escâneres de boa qualidade. Cabe mencionar que é considerada uma boa imagem, quando ao serem ampliadas, conseguimos ler as notícias na tela do computador. Esse processo visa à conservação e preservação da memória de uma época;
- d) **edição dos jornais:** após digitalização, os jornais são editados, sendo as suas imagens clareadas, retirada de brilho e de algumas imperfeições. Para isso, foi necessária a compra de *softwares* especializados que auxiliassem na edição das imagens;
- e) **armazenamento das imagens:** as imagens são armazenadas no computador (em HDs externos) e organizadas em várias pastas (por nome de jornal, ano, mês e número);
- f) **transformação em PDF:** nessa etapa, estamos arquivando os jornais, um por um, em arquivos no formato PDF. A ideia é que, posteriormente, quando esse material for disponibilizado para a comunidade, uma pessoa possa escolher e abrir um arquivo com o jornal que deseja, escolhendo a data, mês e edição que seja de seu interesse.

Por fim, é importante destacar que esse material também ficará disponível nos Arquivos Públicos de Uberaba e Uberlândia, assim como nos laboratórios de pesquisa da UFTM e da UFU. Dessa forma, objetivando amplo acesso ao material, possibilitaremos que o material seja utilizado para pesquisas (sócio)linguísticas e de outras áreas das Ciências Humanas.

Identities dynamics: variation and change in the Spanish of América

I Congreso de la Delegación Argentina de la Asociación de Lingüística y Filología de América Latina (ALFAL) y V Jornadas Internacionales de Filología Hispánica

Palavras Finais

Por tudo que expusemos neste artigo, acreditamos que estudar os jornais de uma dada época é algo inesgotável, pois sempre teremos um ponto diferente para acrescentar. Logo, o resgate desses documentos antigos se justifica na medida em que os jornais:

- (i) são, em sua maioria, únicos, praticamente não possuem muitos exemplares, pois já se perderam ao longo do tempo;
- (ii) abordam temas da vida (nascimento, morte, notícias, pedidos, ordens etc.), e qualquer que seja o assunto, de alguma forma, sempre causam impacto no destinatário;
- (iii) representam escritos que se valorizam ao marcar, mais do que outros textos, a peculiaridade da escritura do jornalista.

Esperamos, assim, ter contribuído para preservação do patrimônio cultural das cidades de Uberaba e Uberlândia e, conseqüentemente, do Brasil, possibilitando que esse material, assim como nossa história, não se perca com o passar do tempo. Além disso, buscamos estimular a realização de outras pesquisas acadêmicas, já que o arquivo de jornais dos séculos XIX e XX das cidades de Uberaba e Uberlândia poderá servir como *corpus* para vários trabalhos acadêmicos, como, por exemplo, trabalhos de final de curso (monografias), artigos, trabalhos de especialização, dissertações ou teses nas mais diversas áreas das Ciências Humanas, sobretudo na Linguística (principalmente, Linguística Histórica, Sociolinguística e Análise do Discurso).

Referências

- Bakhtin, Mikhail [1929]. 2004. *Marxismo e Filosofia da Linguagem*, Traduzido por Michel Lahud e Yara Frateschi Vieira, São Paulo, Editora Annablume/Hucitec.
- Bakhtin, Mikhail. 2000. *Estética da criação verbal*, 3ª ed., São Paulo, Editora Martins Fontes.
- Barbosa, Juliana Bertucci e Talita de Cássia Marine. 2007. O uso das cartas como corpus: uma nova possibilidade às Ciências Humanas. *Anais do V Encontro de Linguística de Corpus*, São Paulo, 5: 01-18.

Identities dinámicas: variación y cambio en el español de América

I Congreso de la Delegación Argentina de la Asociación de Lingüística y Filología de América Latina (ALFAL) y V Jornadas Internacionales de Filología Hispánica

Berlinck, Rosane de Andrade; Juliana Bertucci Barbosa e Talita de Cássia Marine. Reflexões teórico-metodológicas sobre fontes para o estudo histórico da língua. 2008. *Revista da ABRALIN*, v.7, 1: 53-79.

Labov, William. 1972. *Sociolinguistic patterns*, Philadelphia, University of Pennsylvania Press.

Labov, William. 1982. Building on Empirical Foundations. In W. Lehmann and Y. Malkiel (organizers), *Perspectives on Historical Linguistics*, Masterdam/Philadelphia, John Benjamins Publishing Company: 17-92.

Labov, William. 1994. *Principles of Linguistic Change: internal factors*, v. 1, Cambridge, Massachusetts/Oxford, Blackwell Publishers.

Labov, William. 2002. *Principles of Linguistic Change: social factors*, v. 2. Massachusetts/Oxford, Blackwell Publishers.

Marcurschi, Luiz Antônio. 1995. Contextualização e explicitude na relação entre fala e escrita. *Anais língua falada e ensino*, Maceió, UFAL: 27-48.

Marcurschi, Luiz Antônio. 2003. *Análise da conversação*, 5ª ed, Série Princípios, São Paulo, Ática.

Marcurschi, Luiz Antônio. 2005. Gêneros textuais: definição e funcionalidade. Em A. P. Dionísio Angela et al (organizadores), *Gêneros textuais e ensino*, 4ª ed, Rio de Janeiro, Lucerna.

Marcurschi, Luiz Antônio. 2007. *Da fala para a escrita: atividades de recontextualização*, 8ª ed. São Paulo, Cortez.

Marine, Talita de Cássia. 2004. *O binarismo dos pronomes demonstrativos no século XX: este vs. aquele ou esse vs. aquele?* Dissertação de Mestrado, Faculdade de Ciências e Letras, UNESP, Araraquara, Inédita.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.

Weinreich, Uriel; William Labov and Marvin Herzog 1968. Empirical foundations for a theory of language change. In W. Lehmann and Y. Malkiel (organizers), *Directions for historical linguistics: a symposium*, Austin-London, University of Texas Press: 95-199.